

(18)



Europäisches Patentamt

European Patent Office

Office européen des brevets

(11) Publication number:

0 252 685

A2

*attach to*

(12)

## EUROPEAN PATENT APPLICATION

(21) Application number: 87305863.0

(59) Int. Cl.<sup>3</sup>: G 01 N 33/574  
C 12 Q 1/68

(22) Date of filing: 02.07.87

(30) Priority: 09.07.86 US 883687

(43) Date of publication of application:  
13.01.88 Bulletin 88/2(84) Designated Contracting States:  
AT BE CH DE ES FR GB GR IT LI LU NL SE(71) Applicant: The Wistar Corporation  
36th Street at Spruce  
Philadelphia Pennsylvania 19104-4268(US)(72) Inventor: Tsujimoto, Yoshihide  
249 Greene Avenue  
Landsdowne Pennsylvania 19050(US)(72) Inventor: Croce, Carlos M  
2401 Pine Street  
Philadelphia Pennsylvania 19103(US)(74) Representative: Bannerman, David Gardner et al,  
Withers & Rogers 4 Dyer's Buildings Holborn  
London, EC1N 2JT(GB)

(64) Diagnostic methods for detecting lymphomas in humans.

(57) The sequence of the protein coding regions of the *bcl-2* gene are provided as well as bacterial clones which produce the proteins. Assays are provided for detecting a class of B-cell neoplasms associated with a chromosome translocation between chromosomes 14 and 18. This translocation is involved in the majority of cases of human follicular lymphomas. One assay employs an antibody which is immunoreactive with a human protein which is over-expressed due to the chromosome translocation. Another assay involves measurement of the amount of mRNA which hybridizes to the gene proximal to the translocation break-point.

09155327

0202000

## DIAGNOSTIC METHODS FOR DETECTING LYMPHOMAS IN HUMANS

This invention was made with government support under Grant CA 39860 from the National Cancer Institute. The United States Government has certain rights in this invention.

### BACKGROUND OF THE INVENTION

Specific chromosomal rearrangements, predominantly translocations and inversions, are observed in the great majority of human hematopoietic malignancies. In Burkitt lymphoma the specific chromosomal translocations result in the juxtaposition of one of the three human immunoglobulin loci and the c-myc oncogene. In diffuse B-cell lymphomas, multiple myelomas and chronic lymphocytic leukemias of the B-cell type carrying the t(11;14) (q13;q32) chromosome translocation, the bcl-1 locus is translocated to the heavy-chain locus on chromosome 14. In most cases of follicular lymphoma, one of the most common human hematopoietic malignancies, a t(14;18) (q32;q21) chromosome translocation has been observed. This translocation moves the bcl-2 gene to a position adjacent to the heavy-chain locus. In one cell line derived from a leukemic patient having both a t(14;8) and a t(14;18) translocation enhanced mRNA production from the bcl-2 gene was observed. (Tsujimoto et al, Science, Vol. 228, pages 1440-1443 (1985).) It was concluded there that the transcription unit of the

bcl-2 gene spans the chromosome break-point, and thus the oncogene protein is likely to be structurally altered in the B-cell neoplasms. Surprisingly, it has now been found that the translocation does not alter the oncogene protein itself, as the translocation break-points occur downstream from the actual protein coding sequences. Thus oncogenesis may be solely due to the overproduction of the normal human gene products of the bcl-2 gene.

Effective treatment for cancer is often dependent upon an early and proper diagnosis of the malignancy. There is thus a need for simple and accurate diagnostic methods for detecting and identifying human malignancies, such as follicular lymphomas, in general, and follicular lymphomas in particular.

#### SUMMARY OF THE INVENTION

It is an object of the present invention to provide a diagnostic method for detecting B-cell neoplasms having t(14;18) translocations in a human.

It is another object of the present invention to provide a human bcl-2 gene which can be expressed in bacteria.

It is yet another object of the present invention to provide a substantially pure protein preparation of a bcl-2 gene product.

It is still another object of the present invention to provide a diagnostic method of detecting B-cell neoplasms employing an antibody which is immunoreactive with a bcl-2 gene product.

These and other objects of the invention are achieved by one or more of the following embodiments.

In one embodiment a diagnostic method for detecting B-cell neoplasms in a human is provided, comprising the steps of:

isolating B-cells from the human;  
extracting proteins from said B-cells to form a test sample;

contacting said test sample with an antibody which is immunoreactive with a gene product of the bcl-2 gene, under conditions where antibody-antigen complexes are formed and are stable;

quantitating the amount of antibody-antigen complexes formed with said test sample; and

comparing the amount of antibody-antigen complexes formed with said test sample to the amount formed with a control sample of proteins, a ratio of test sample complexes to control sample complexes of greater than about ten indicating B-cell neoplasm, said control sample of proteins extracted from cells selected from the group consisting of: B-cells from a normal human, cells from an established normal B-cell or pre-B-cell line, and non-B-cells from said human.

Also provided are forms of the human bcl-2 gene which are substantially free of introns. Such genes can be replicated and expressed in bacteria to form proteins having the same primary structure as the bcl-2 proteins produced in humans.

Also provided by the present invention are substantially pure preparations of proteins having an N-terminal end encoded by the first exon of the human bcl-2 gene.

In yet another embodiment of the present invention a diagnostic method for detecting B-cell neoplasms in a human is provided comprising the steps of:

isolating B-cells from the humans;  
extracting RNA from said B-cells to form a test sample;

contacting said test sample with a DNA probe containing a sequence of at least 15 neucleotides in length derived from the human bcl-2 gene, under conditions where homologous RNA-DNA hybrids form and are stable;

quantitating the amount of RNA-DNA hybrids formed with the test sample; and

comparing the amount of RNA-DNA hybrids formed with said test sample to the amount formed with a control sample of RNA, a ratio of test sample hybrids to control sample hybrids of greater than about ten indicating B-cell neoplasm, said control sample of RNA extracted from cells selected from the group consisting of: B-cells from a normal human, cells from a normal B-cell or pre-B-cell line, and non-B-cells from said human.

#### BRIEF DESCRIPTION OF THE FIGURES

Figure 1 shows a genomic restriction map of chromosome 18, and the structure of cDNA clones.

Figure 2 shows the nucleotide sequence of bcl-2 cDNA corresponding to the 5.5 kb transcript. Only the sequences surrounding the open reading frame are shown.

Figure 3 shows the nucleotide sequence of bcl-2 cDNA corresponding to the 3.5 kb transcript. Only the sequences surrounding the open reading frame are shown.

#### DETAILED DESCRIPTION OF THE INVENTION

According to one aspect of the present invention it has been discovered that B-cell neoplasms which are associated with t(14;18)

chromosome translocations cause an increase in the expression of both the mRNA and the protein products of the bcl-2 gene. The expression in the neoplastic B-cells is generally about ten-fold higher than the amount expressed by normal cells. This elevated expression can be used as a diagnostic tool to detect the B-cell neoplasms in humans. Such neoplasms include follicular lymphomas as well as other lymphomas.

It has been found that three species of mRNA are transcribed from the bcl-2 gene. At least two different protein products have been identified. A 239 amino acid protein, designated bcl-2-alpha is translated from a 5.5 kb mRNA. A protein of 205 amino acids, bcl-2-beta is translated from a 3.5 kb mRNA. It is thought that bcl-2-alpha is also translated from the 8.5 kb mRNA species. All three size species of transcript share substantial sequence homology in the 5' portion of the gene, termed the first exon or 5' exon. The larger two transcripts appear to be spliced to a second exon which is at least 50 kb distal to the first. The splice site is in the midst of the protein coding sequence. Thus, the 3.5 kb transcript codes for a protein having a different carboxyl terminus than the proteins coded by the two larger sized transcripts.

The hotspot for chromosome break-points among the t(14;18) translocations of follicular lymphomas map 3' to the protein coding region. Therefore it has now been found that the translocations do not alter the primary structure of the protein products.

Bacterial isolates available from the ATCC, under Accession Numbers 67147 and 67148, can be used to express bcl-2 gene products alpha and beta, respectively, in bacteria. The clones of the bcl-2 gene were obtained via cDNA cloning and so do not contain

introns. Thus these clones can be expressed in bacteria to make products having the same primary sequence as those made in the human body. After growing the bacteria under suitable conditions which are well known in the art, the cells can be harvested and disrupted to extract total cellular protein. The protein can then, for example, be placed on a sizing column such as Sepharose™ or agarose beads, and proteins of the correct molecular weight, i.e., between about 20 and 30 kD can be collected.

Further purification can be effected by use of an anti-bcl-2 antibody. Such an antibody can be used to immunoprecipitate bcl-2 proteins from the set of cellular proteins of the correct approximate molecular weight. Such antibodies can, for example, be raised against polypeptides synthesized according to the sequence shown in Figures 2 and 3. Alternatively, the antibodies can be raised against fusion proteins, which contain bcl-2 sequences as well as those of another protein. An example of an antibody raised to a fusion protein is discussed, infra. After immunoprecipitation, the bcl-2 proteins can be released from the antibodies to provide a substantially pure preparation of bcl-2 proteins.

If it is desired that bcl-2-alpha (approximately 26 kD) be separated from bcl-2-beta (approximately 22 kD) this separation can be achieved using either polyacrylamide gels or additional sizing or gel filtration columns. Of course, other separations are possible based on the inherent differences between the two proteins at their carboxyl termini. Techniques for using gel filtration columns and immunoprecipitation and antibody releasing are all well known in the art.

Any source of B-cells is suitable for use in the diagnostic test of this invention. B-cells can be isolated from the lymph nodes. Alternatively, the diagnostic test may employ a sample of peripheral blood obtained from an individual who is being screened for the presence of a B-cell neoplasm. Means for separating B-cells from the peripheral blood are well known in the art. For example, erythrocytes and granulocytes may be separated from the B-cells by centrifugation in a liquid having a density intermediate between the groups of cells to be separated.

Extraction of proteins from B-cells may be performed by any of the many means known in the art. For example, cells may be lysed by a detergent or by mechanical means. If desired, nucleic acids can be removed from the cell preparation by enzymatic digestion or by precipitation with agents such as streptomycin. Once again, such means are well known in the art.

Antibodies can be generated which are immunoreactive with the bcl-2 proteins by immunization of animals with a fusion protein consisting of a portion of the beta-galactosidase protein of E.coli and a portion of the human bcl-2 proteins. Preferably, the bcl-2 portion will contain sequences which are common to both bcl-2-alpha and bcl-2-beta. If desired, such a fusion protein can be purified using the properties which it shares with beta-galactosidase. Anti-sera raised against such a fusion protein in rabbits have been found to be immunoreactive with both bcl-2-alpha and bcl-2-beta in vitro. Furthermore, using this anti-sera in immunofluorescent techniques it is possible to determine cellular location of bcl-2 proteins in cells which have been fixed.



Antibodies can also be produced by immunization of animals, such as mice, rabbits and the like, with bcl-2-alpha, bcl-2-beta, fragments of them, or both. Alternatively, monoclonal antibodies can be generated using immortalized cell lines to provide uniform and continual antibody sources. Techniques for generating such antibodies are well known in the art. Appropriate antibodies can be screened using the natural gene products of bcl-2 or the fusion protein discussed above. While it is preferred that the antibody used in the diagnostic method immunoreact with both bcl-2-alpha and beta, an antibody may successfully be used which immunoreacts with only one of them.

The extracted proteins from the B-cells may be contacted with the antibody under suitable conditions for antibody-antigen complex formation. Generally, such conditions are physiological conditions. The protein extract may be bound to a solid support such as a nitrocellulose filter or a microtiter plate.

The antibody will generally bear a "tag" such as a radiolabel, a fluorescent label or an enzyme conjugate which under appropriate conditions produces a colored reaction product. Once again, such "tags" are quite well known in the art. Alternatively, if the antibody is not tagged, it can be detected by means of a second antibody from another species which is reacted with the first antibody. Of course, it is preferred for means of this diagnostic method that the immunological technique be as quantitatively sensitive as possible.

Means of detection of the antibody-antigen complexes will depend upon the method of tagging used for the antibody. For example, radiolabel tags can be detected by autoradiography or scintillation counting, while the products of enzyme-linked antibodies can be detected spectrophotometrically.

A parallel sample to the test sample is employed to provide the control. The control sample consists of an equivalent amount of proteins extracted from cells, preferably in the same manner as those of the test sample. The amount of protein can readily be determined employing techniques well known in the art, such as the Lowry or Bradford techniques. The cells used for preparing the control sample may be selected from the group consisting of B-cells from a normal human, cells from an established normal B-cell or pre-B-cell line, and non-B-cells from the human who is being screened for the neoplasm.

It is a finding of the present invention that in cases where a translocation has occurred between chromosomes 14 and 18 the level of bcl-2 protein detected immunologically in the B-cells is at least 10-fold higher than the amount of bcl-2 protein found in normal B-cells, in pre-B-cells, or in other non-B-cells from the same human.

To screen for elevated levels of mRNA transcribed from the bcl-2 gene, again one must isolate B-cells from the human who is to be screened. Any of the many methods known in the art are suitable. Total RNA extracted from the B-cells may be used, or alternatively mRNA may be isolated from the total cellular RNA. The mRNA may be purified, for example, by affinity chromatography on oligo(dT) cellulose which binds to the poly(A) tract at the 3' end of most mRNA. As is well known to those of skill in the art, it is essential that ribonuclease activity be minimized during preparation and assaying.

A DNA probe may be selected from any of the protein coding sequences of the bcl-2 gene. Preferably, the probe will be selected from sequences of the 5' or first exon of the gene, so that all three

species of RNA can be detected. Alternatively the probe can be selected from sequences which hybridize exclusively with the 3.5 kb transcript or only with the 5.5 kb and the 8.5 kb transcript. Preferably the probe contains at least 15 nucleotides of the bcl-2 sequence. Suitable plasmid molecules which may be used as probes have been deposited at the ATCC under Deposit Numbers 67147 and 67148. Of course, other suitable probes may be synthesized or derived from these or other bcl-2 sequences. In order to perform the hybridization it is desirable that the probe be single stranded. Thus if the probe is double stranded, it should be denatured to single stranded form. Means for denaturing are well known in the art, including alkali or heat treatment. The probe can then be contacted with the RNA derived from the B-cells under conditions where homologous RNA-DNA hybrids form and are stable. Such conditions are well known in the art. Means for detecting hybrids are many and well known, but often involves use of radiolabeled probes and nucleases which degrade single stranded DNA. Other methods may be used.

Control samples can be derived from any of the cell sources described above for use in the antibody diagnostic test. Samples and control should be prepared in parallel under similar conditions. If comparison of the test and control sample hybridization shows a greater than about a ten-fold excess in the test sample, a B-cell neoplasm is indicated.

The following examples do not limit the scope of the invention but are merely illustrative.

#### EXAMPLE 1

A cDNA library from polyA<sup>+</sup> mRNA of the pre-B-cell leukemia cell line 380 was constructed. Cytoplasmic RNA was extracted by the procedure described in ar-Rushdi, et al (1982) Somatic Cell Genetics, Vol. 8, pp. 151-161. PolyA<sup>+</sup> RNA was selected by oligo(dT) column chromatography as described in Aviv and Leder, (1972) Proceedings of National Academy of Sciences, USA, Vol. 69, pp. 1408-1412. Double stranded cDNA was synthesized from mRNA by reverse transcriptase (Life Science, Inc., Florida) using oligo(dT) as primer as described in Maniatis et al (1982) Molecular Cloning, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York. After EcoRI linker ligation, the double stranded cDNA was cloned into lambda gt 11 phage vectors, Young and Davis (1983) Proceedings of the National Academy of Sciences, USA, Vol. 80, pp. 1194-1198. By screening approximately  $2 \times 10^5$  recombinant clones with a DNA probe consisting of a segment of chromosome 18 which spans the hotspot of break-points of the translocation of chromosome 18 to chromosome 14, three independent cDNA clones were obtained which are overlapping (B3, B4, and B10). As shown in Figure 1, clone B3 contains nineteen A residues at the end, indicating that this clone represents the 3' end of the mRNA. The restriction maps of the cDNA clones and genomic sequences are colinear from the 3' end of cDNA clone B3 until just before the BamHI site of the cDNA sequence. The cDNA sequence just beyond this point diverges from the genomic sequence. Thus, the cDNA sequences consist of at least two genomic regions.

When the 5' part of cDNA clone B4 (5' end to the BamHI site) is used to probe the cDNA bank, another set of clones is obtained,

including clone B15 and clone B16 (see Fig. 1). These two cDNA clones have the same sequences at the 5' region as does clone B4, but they have totally different sequences at the 3' region than do clones B3, B4, and B10. Thus cDNA cloning has yielded two different sets of clones, indicating that the bcl-2 gene is transcribed into at least two different mRNAs.

In order to obtain cDNA sequences further upstream (in the 5' direction), a cDNA library was constructed using the primer extension method. An oligonucleotide (15-mer) was synthesized and used as a primer for reverse transcriptase as described in Maniatis, supra, and in Gubler and Hoffman (1983) *Gene*, Vol. 25, pp. 263-269. Three clones were obtained by this method, clone B6-3, B22-1, and B9.

#### EXAMPLE II

Two different probes were used to visualize the mRNA species corresponding to the bcl-2 gene in Northern blot hybridizations. The first probe (probe A in Figure 1) contains genomic DNA of chromosome 18 which spans the break-point hotspot and also corresponds to the 3' exon. The other probe used was cDNA clone B22-1 which corresponds to the 5' exon (the first exon). RNA was glyoxalated, run on 1% agarose gels, and blotted to nitrocellulose filters as described in Thomas, *Proceedings of the National Academy of Sciences, USA* (1980) Vol. 77, pp. 5201-5205. The nitrocellulose filter was hybridized with <sup>32</sup>P-labelled probe in 50% formamide, 4 X SSC, 0.1% SDS at 37°C and finally washed with 2 X SSC (0.3M NaCl, 0.03 M Na-citrate, pH 7) at 65°C.

The genomic DNA probe A detected two transcripts, 8.5 kb and 5.5 kb in length. The cDNA probe B22-1 detected the same transcripts as probe A, as well as an additional transcript of 3.5 kb.

The 8.5 kb mRNA was also shown to hybridize to a genomic DNA probe from chromosome 18 which is 3' to genomic probe A (and is indicated in Fig. 1 as probe B).

These data indicate that the bcl-2 gene is transcribed into three ~~mRNAs~~ of different sizes. The possibility that these mRNAs are derived from different but related genes is excluded by the fact that under the same hybridization conditions which were used for the Northern blot hybridization, these probes detect only one cellular gene.

#### EXAMPLE III

The nucleotide sequence of overlapping cDNA clones was determined by Maxam and Gilbert's chemical degradation method, (Proceedings of the National Academy of Sciences, USA, Vol. 74, pp. 560-564 (1977), or Sanger's chain termination method, Proceedings of the National Academy of Sciences, USA, Vol. 74, pp. 5463-5467 (1977). Both strands of DNA were sequenced. The nucleotide sequence derived from the 5.5 kb transcript is shown in Figure 2. The DNA sequence of 5105 base pairs (bp) reveals one possible open reading frame consisting of 239 amino acid residues (bcl-2-alpha). The nucleotide sequence corresponding to the 3.5 kb transcript is shown in Figure 3. This transcript codes for a protein consisting of 205 amino acid residues (bcl-2-beta), which differs from the bcl-2-alpha protein at the carboxyl terminus.

**CLAIMS**

1. A diagnostic method for detecting B-cell neoplasms in a human, comprising the steps of:

isolating B-cells from the human;

extracting proteins from said B-cells to form a test sample;

contacting said test sample with an antibody which is immunoreactive with a gene product of the bcl-2 gene under conditions where antibody-antigen complexes form and are stable to form antibody-antigen complexes;

quantitating the amount of antibody-antigen complexes formed with said test sample;

comparing the amount of antibody-antigen complexes formed with said test sample to the amount formed with a control sample of proteins, a ratio of test sample complexes to control sample complexes of greater than about ten indicating B-cell neoplasm, said control sample of proteins extracted from cells selected from the group consisting of: B-cells from a normal human, cells from an established normal B-cell or pre-B-cell line, and non-B-cells from said human.

2. The method of claim 1 wherein the antibody is radiolabeled and the step of quantitating occurs by measuring the amount of radioactivity in antibody-antigen complexes.

3. The method of claim 1 wherein the antibody is conjugated to a fluorescent molecule and the step of quantitating is performed by measuring the amount of fluorescence emitted by said antibody-antigen complexes.

- 15 -

4. The method of claim 1 wherein said step of quantitating is performed by means of use of a second antibody immunoreactive with said antibody which is immunoreactive with a gene product of bcl-2 gene.

5. The method of claim 1 wherein the antibody is immunoreactive with bcl-2-alpha.

6. The method of claim 1 wherein the antibody is immunoreactive with bcl-2-beta.

7. The method of claim 1 wherein the antibody is immunoreactive with both bcl-2-alpha and bcl-2-beta.

8. A human bcl-2 gene substantially free of introns.

9. The gene of claim 8 wherein said gene is replicated in bacteria.

10. The gene of claim 9 wherein said gene is expressed in bacteria.

11. A substantially pure preparation of a protein having an N-terminal end encoded by the first exon of the human bcl-2 gene.

12. The protein preparation of claim 11 wherein said protein is bcl-2-alpha having about 239 amino acid residues.

13. The protein preparation of claim 11 wherein said protein is bcl-2-beta having about 205 amino acid residues.

14. A diagnostic method for detecting B-cell neoplasms in a human comprising the steps of:

isolating B-cells from the human;

extracting RNA from said B-cells to form a test sample;

contacting said test sample with a DNA probe containing a sequence of at least 15 nucleotides in length derived



from the human bcl-2 gene under conditions where homologous RNA-DNA hybrids form and are stable;

quantitating the amount of RNA-DNA hybrids formed in the test sample; and

comparing the amount of RNA-DNA hybrids formed with said test sample to the amount formed with a control sample of RNA, a ratio of test sample hybrids to control sample hybrids of greater than about ten indicating B-cell neoplasm, said control sample of RNA extracted from cells selected from the group consisting of: B-cells from a normal human, cells from a normal B-cell or pre-B-cell line, and non-B-cells from said human.

15. The method of claim 14 wherein said DNA probe hybridizes with the 8.5 kb, the 5.5 kb, and the 3.5 kb bcl-2 transcript.

16. The method of claim 14 wherein said DNA probe hybridizes with the 8.5 kb and the 5.5 kb bcl-2 transcript.

17. The method of claim 14 wherein said DNA probe hybridizes exclusively with the 3.5 kb bcl-2 transcript.

Not eingereicht / Newly filed  
Nouvellement déposé

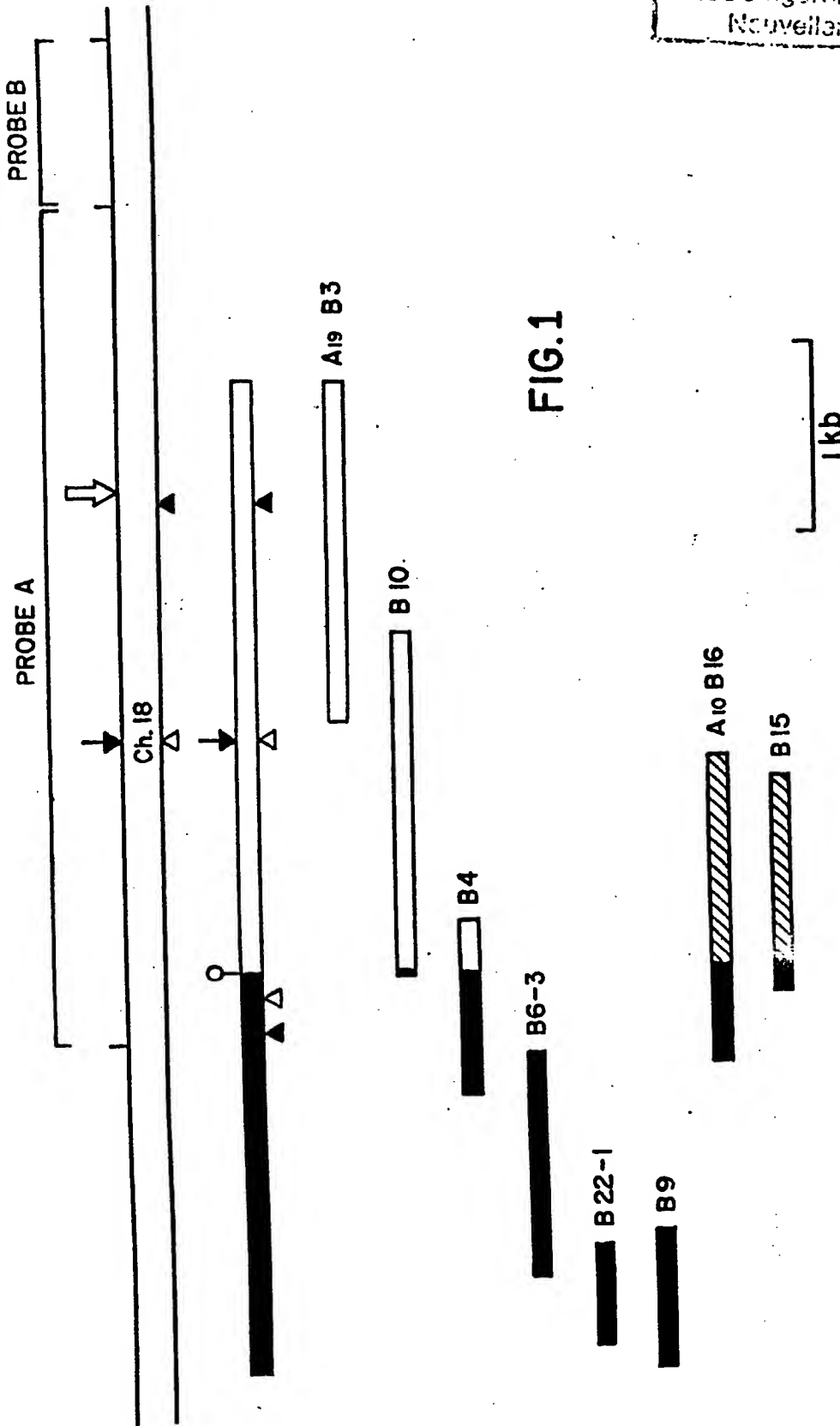


FIG.1

## LEGEND

- ↓ = HOT SPOT OF THE BREAKPOINT OF t(14;18) TRANSLOCATIONS.
- ▭ = SECOND EXON  
 ▨ = FIRST EXON  
 ▩ = FIRST EXON
- | = Hind III  
 ○ = Bam HI  
 ▲ = Sst I
- ↓ = Eco RI

Revised 10/18/1987  
Revised 10/18/1987

CCCTCCGCGC CGCCTGCCCG CCCGCCCGCC GCGCTCCCGC CCGCCGCTCT -1401  
CCGTGGCCCC GCGCGGCTGC CGCCGCCGCC GCTGCCAGCG AAGGTGCCCG  
GGCTCCGGGC CCTCCCTGCC GCGCGCCGTC AGCGCTCGGA CGGAAGTCCG  
CGACGGGAGG TCCGGGAGGC GACCGTAGTC GCGCCGCGCG CCAGGACCAG  
GAGGAGGAGA AAGGGTGGCG AGCCCGGAGG CCGGGTCCGC CGGTGGGGTG  
CAGCGGAAGA GGGGGTCCAG GGGGGAGAAC TTCGTAGCAG TCATCCTTTT  
TAGGAAAAGA GGGAAAAAAT AAAACCCTCC CCCACCACCT CCTTCTCCCC  
ACCCCTCGCC GCACCACACA CAGCGCGGGC TTCTAGCGCT CGGCACCGGC  
GGGCCAGGCG CGTCCTGCCT TCATTTATCC AGCAGCTTTT CGGAAAATGC -1001

ATTTGCTGTT CGGAGTTTAA TCAGAAGACG ATTCCTGCCT CCGTCCCCGG  
CTCCTTCATC GTCCCATCTC CCCTGTCTCT CTCCTGGGGA GGCGTGAAGC  
GGTCCCGTGG ATAGAGATTC ATGCCTGTGT CCGCGCGTGT GTGCGCGCGT  
ATAAATTGCC GAGAAGGGGA AAACATCACA GGAATTCTGC GAATACCGGA  
CTGAAAATTG TAATTCATCT GCCGCCGCCG CTGCCAAAAA AAAACTCGAG  
CTCTTGAGAT CTCCGGTTGG GATTCCTGCG GATTGACATT TCTGTGAAGC  
AGAAGTCTGG GAATCGATCT GGAAATCCTC CTAATTTTAA CTCCCTCTCC  
CCCCGACTCC TGATTCATTG GGAAGTTTCA AATCAGCTAT AACTGGAGAG  
TGCTGAAGAT TGATGGGATC GTTGCCTTAT GCATTTGTTT TGGTTTTACA  
AAAAGGAAAC TTGACAGAGG ATCATGCTGT ACTTAAAAAA TACAAGTAAG -501

TCTCGCACAG GAAATTGGTT TAATGTAACT TTCAATGGAA ACCTTTGAGA  
TTTTTTACTT AAAGTGCATT CGAGTAAATT TAATTTCCAG GCAGCTTAAT  
ACATTGTTTT TAGCCGTGTT ACTTGTAAGT TGTATGCCCT GCTTTCACCT  
AGTGTGTACA GGGAAACGCA CCTGATTTTT TACTTATTAG TTTGTTTTTT  
CTTTAACCTT TCAGCATCAC AGAGGAAGTA GACTGATATT AACAATACTT  
ACTAATAATA ACGTGCCTCA TGAAATAAAG ATCCGAAAGG AATTGGAATA  
AAAATTTTCT GCGTCTCATG CCAAGAGGGA AACACCAGAA TCAAGTGTTT  
CGCGTGATTG AAGACACCCC CTCGTCCAAG AATGCAAAGC ACATCCAATA  
AAATAGCTGG ATTATAACTC CTCTTCTTTC TCTGGGGGCC GTGGGGTGGG  
AGCTGGGGCG AGAGGTGCCG TTGGCCCCCG TTGCTTTTCC TCTGGGAAGG -1

FIG. 2A

1 MET ALA HIS ALA GLY ARG THR GLY TYR ASP ASN ARG GLU ILE VAL  
 ATG GCG CAC GCT GGG AGA ACG GGG TAC GAC AAC CGG CAG ATA GTG  
  
 MET LYS TYR ILE HIS TYR LYS LEU SER GLN ARG GLY TYR GLU TRP  
 46 ATG AAG TAC ATC CAT TAT AAG CTG TCG CAG AGG GGC TAC GAG TGG  
  
 ASP ALA GLY ASP VAL GLY ALA ALA PRO PRO GLY ALA ALA PRO ALA  
 91 GAT GCG GGA GAT GTG GGC GCC GCG CCC CCG GGG GCC GCC CCC GCA  
  
 PRO GLY ILE PHE SER SER GLN PRO GLY HIS THR PRO HIS PRO ALA  
 136 CCG GGC ATC TTC TCC TCC CAG CCC GGG CAC ACG CCC CAT CCA GCC  
  
 ALA SER ARG ASP PRO VAL ALA ARG THR SER PRO LEU GLN THR PRO  
 181 GCA TCC CGC GAC CCG GTC GCC AGG ACG TCG CCG CTG CAG ACC CCG  
  
 ALA ALA PRO GLY ALA ALA ALA GLY PRO ALA LEU SER PRO VAL PRO  
 226 GCT GCC CCC GGC GCC GCC GCG GGG CCT GCG CTC AGC CCG GTG CCA  
  
 PRO VAL VAL HIS LEU ALA LEU ARG GLN ALA GLY ASP ASP PHE SER  
 271 CCT GTG GTC CAC CTG GCC CTC CGC CAA GCC GGC GAC GAC TTC TCC  
  
 ARG ARG TYR ARG GLY ASP PHE ALA GLU MET SER SER GLN LEU HIS  
 316 CGC CGC TAC CGC GGC GAC TTC GCC GAG ATG TCC AGC CAG CTG CAC  
  
 LEU THR PRO PHE THR ALA ARG GLY ARG PHE ALA THR VAL VAL GLU  
 361 CTG ACG CCC TTC ACC GCG CGG GGA CGC TTT GCC ACG GTG GTG GAG  
  
 GLU LEU PHE ARG ASP GLY VAL ASN TRP GLY ARG ILE VAL ALA PHE  
 406 GAG CTC TTC AGG GAC GGG GTG AAC TGG GGG AGG ATT GTG GCC TTC  
  
 PHE GLU PHE GLY GLY VAL MET CYS VAL GLU SER VAL ASN ARG GLU  
 451 TTT GAG TTC GGT GGG GTC ATG TGT GTG GAG AGC GTC AAC CGG GAG  
  
 MET SER PRO LEU VAL ASP ASN ILE ALA LEU TRP MET THR GLU TYR  
 496 ATG TCG CCC CTG GTG GAC AAC ATC GCC CTG TGG ATG ACT GAG TAC  
  
 LEU ASN ARG HIS LEU HIS THR TRP ILE GLN ASP ASN GLY GLY TRP  
 541 CTG AAC CGG CAC CTG CAC ACC TGG ATC CAG GAT AAC GGA GGC TGG  
  
 ASP ALA PHE VAL GLU LEU TYR GLY PRO SER MET ARG PRO LEU PHE  
 586 GAT GCC TTT GTG GAA CTG TAC GGC CCC AGC ATG CCG CCT CTG TTT  
  
 ASP PHE SER TRP LEU SER LEU LYS THR LEU LEU SER LEU ALA LEU  
 631 GAT TTC TCC TGG CTG TCT CTG AAG ACT CTG CTC AGT TTG GCC CTG  
  
 VAL GLY ALA CYS ILE THR LEU GLY ALA TYR LEU SER HIS LYS  
 676 GTG GGA GCT TGC ATC ACC CTG GGT GCC TAT CTG AGC CAC AAG TGA

FIG. 2B

Neu eingereicht / Newly filed  
Nouvellement déposé

AGTCAACATG CCTGCCCCAA ACAAATATGC  
AAAAGGTTCA CTAAAGCAGT AGAAATAATA TGCATTGTCA GTGATGTACC  
ATGAAACAAA GCTGCAGGCT GTTTAAGAAA AAATAACACA CATATAAACA  
TCACACACAC AGACAGACAC ACACACACAC AACAATTAAC AGTCTTCAGG  
CAAAACGTCC AATCAGCTAT TTAAGGCCAA AGGGAAATAT CATTATTTTT  
TTACATTATT AAGAAAAAAG ATTTATTTAT TTAAGACAGT CCCATCAAAA

1001 CTCCGTCTTT GGAAATCCGA CCACTAATTG CCAAACACCG CTTCGTGTGG  
CTCCACCTGG ATGTTCTGTG CCTGTAAACA TAGATTGCGT TTCCATGTTG  
TTGGCCGGAT CACCATCTGA AGAGCAGACG GATGGA AAAA GGACCTGATC  
ATTGGGGAAG CTGGCTTTCT GGCTGCTGGA GGCTGGGGAG AAGGTGTTCA  
TTCACTTGCA TTTCTTTGCC CTGGGGGCGT GATATTAACA GAGGGAGCGT  
TCCCGTGGGG GGAAGTCCAT GCCTCCCTGG CCTGAAGAAG AGACTCTTTG  
CATATGACTC ACATGATGCA TACCTGGTGG GAGGAAAAGA GTTGGGAACT  
TCAGATGGAC CTAGTACCCA CTGAGATTTC CACGCCGAAG GACAGCGATG  
GGAAAAATGC CCTTAAATCA TAGGAAAGTA TTTTTTTAAG CTACCAATTG  
TGCCGAGAAA AGCATTTTAG CAATTTATAC AATATCATCC AGTACCTTAA

1501 ACCCTGATTG TGTATATTCA TATATTTTGG ATACGCACCC CCCAACTCCC  
AATACTGGCT CTGTCTGAGT AAGAAACAGA ATCCTCTGGA ACTTGAGGAA  
GTGAACATTT CGGTGACTTC CGATCAGGAA GGCTAGAGTT ACCCAGAGCA  
TCAGGCCGCC ACAAGTGCCCT GCTTTTAGGA GACCGAAGTC CGCAGAACCT  
ACCTGTGTCC CAGCTTGAGG GCCTGGTCTT GGAAGTGAAG CGGGCCCTCA  
CTGGCCTCCT CCAGGGATGA TCAACAGGGT AGTGTGGTCT CGGAATGTCT  
GGAAGCTGAT GGATGGAGCT CAGAATTCCA CTGTCAAGAA AGAGCAGTAG  
AGGGGTGTGG CTGGGCCTGT CACCCTGGGG CCCTCCAGGT AGGCCCGTTT  
TCACGTGGAG CATAGGAGCC ACGACCCTTC TTAAGACATG TATCACTGTA  
GAGGGAAGGA ACAGAGGCCC TGGGCCTTCC TATCAGAAGG ACATGGTGAA

2001 GGCTGGGAAC GTGAGGAGAG GCAATGGCCA CGGCCCATTT TGGCTGTAGC  
ACATGGCAGG TTGGCTGTGT GGCCTTGCC ACCTGTGAGT TTAAAGCAAG  
GCTTTAAATG ACTTTGGAGA GGGTCACAAA TCCTAAAAGA AGCATTGAAG  
TGAGGTGTCA TGGATTAATT GACCCCTGTC TATGGAATTA CATGTAAAAC  
ATTATCTTGT CACTGTAGTT TGGTTTTATT TGAAAACCTG AAAAAA AAAA  
AGTTCCAGGT GTGGAATATG GGGGTTATCT GTACATCCTG GGGCATTAAA  
AAAAAATCAA TGGTGGGGAA CTATAAAGAA GTAACAAAAG AAGTGACATC  
TTCAGCAAAT AAAGTAGGAA ATTTTTTTTT CTTCAGTTT AGAATCAGCC  
TTGAAACATT GATGGAATAA CTCTGTGGCA TTATTGCATT ATATACCATT  
TATCTGTATT AACTTTGGAA TGTACTCTGT TCAATGTTTA ATGCTGTGGT

FIG. 2C

Not eingereicht / Noty filed  
 Abgleichung / Alignment

2501 TGATATTTTCG AAAGCTGCTT TAAAAAATA CATGCATCTC AGCGTTTTTTT  
 TGTTTTTAAT TGTATTTAGT TATGGCCTAT ACACTATTTG TGAGCAAAGG  
 TGATCGTTTT CTGTTTGAGA TTTTATCTC TTGATTCTTC AAAAGCATTC  
 TGAGAAGGTG AGATAAGCCC TGAGTCTCAG CTACCTAAGA AAAACCTGGA  
 TGTCACTGGC CACTGAGGAG CTTTGTTCAT ACCAAGTCAT GTGCATTTCC  
 ACGTCAACAG AATTGTTTAT TGTGACAGTT ATATCTGTTG TCCCTTTGAC  
 CTTGTTTCTT GAAGGTTTCC TCGTCCCTGG GCAATTCCGC ATTTAATTCA  
 TGGTATTCAG GATTACATGC ATGTTTGTTT AAACCCATGA GATTCATTCA  
 GTTAAAAATC CAGATGGCGA ATGACCAGCA GATTCAAATC TATGGTGTTT  
 TGACCTTTAG AGAGTTGCTT TACGTGGCCT GTTCAACAC AGACCCACCC

3001 AGAGCCCTCC TGGCCTCCTT CCGGGGGGGC TTTCTCATGG CTGTCCTTCA  
 GGGTCTTCCT GAAATGCAGT GGTGCTTACG CTCCACCAAG AAAGCAGGAA  
 ACCTGTGGTA TGAAGCCAGA CCTCCCCGGC GGGCCTCAGG GAACAGAATG  
 ATCAGACCTT TGAATGATTC TAATTTTTAA GCAAAATATT ATTTTATGAA  
 AGGTTTACAT TGTCAAAGTG ATGAATATGG AATATCCAAT CCTGTGCTGC  
 TATCCTGCCA AAATCATTTT AATGGAGTCA GTTTGCAGTA TGCTCCACGT  
 GGTAAGATCC TCCAAGCTGC TTTAGAAGTA ACAATGAAGA ACGTGGACGT  
 TTTTAATATA AAGCCTGTTT TGTCTTTTGT TGTGTGTTCA ACGGGATTCA  
 CAGAGTATTT GAAAAATGTA TATATATTAA GAGGTCACGG GGGCTAATTG  
 CTAGCTGGCT GCCTTTTGCT GTGGGGTTTT GTTACCTGGT TTTAATAACA

3501 GTAAATGTGC CCAGCCTCTT GGCCCCAGAA CTGTACAGTA TTGTGGCTGC  
 ACTTGCTCTA AGAGTAGTTG ATGTTGCATT TTCCTTATTG TTA AAAACAT  
 GTTAGAAGCA ATGAATGTAT ATAAAAGCAA AAAAAAAAAA AAAAAAA

FIG. 2D

Not a true 100% match  
 Novelty of the sequence

TGATTGAAGACACCCCCCTGTCACAAGAATGCAAAGCACATCCAATAAAATAGCTGGATTATAACTOCTCTTCTTTCTCTG

+1

GGGGCCGTGGGGTGGGAGCTGGGGCGAGAGGTGCGGTGGCCCCCGTTGCTTTTCTCTGCGGAAGG MET ALA HIS  
 ATG GCG CAC

ALA GLY ARG THR GLY TYR ASP ASN ARG GLU ILE VAL MET LYS TYR ILE HIS TYR LYS LEU  
 GCT GGG AGA ACG GCG TAC GAC AAC CCG GAG ATA GTG ATG AAG TAC ATC CAT TAT AAG CTG

100

SER GLN ARG GLY TYR GLU TRP ASP ALA GLY ASP VAL GLY ALA ALA PRO PRO GLY ALA ALA  
 TCG CAG AGG GCG TAC GAG TGG GAT GCG GGA GAT GTG GCG GCG GCG CCG CCG GCG GCG GCG

PRO ALA PRO GLY ILE PHE SER SER GLN PRO GLY HIS THR PRO HIS PRO ALA ALA SER ARG  
 CCG GCA CCG GCG ATC TTC TOC TOC CAG CCG CCG CAC ACG CCG CAT CCA GCG GCA TOC CCG

200

ASP PRO VAL ALA ARG THR SER PRO LEU GLN THR PRO ALA ALA PRO GLY ALA ALA ALA GLY  
 GAC CCG GTC GCG AGG ACG TOG CCG CTG CAG ACG CCG GCT GCG CCG GCG GCG GCG GCG

300

PRO ALA LEU SER PRO VAL PRO PRO VAL VAL HIS LEU ALA LEU ARG GLN ALA GLY ASP ASP  
 OCT GCG CTC ACG CCG GTG CCA OCT GTG GTC CAC CTG GCG CTC CCG CAA GCG GCG GAC GAC

PHE SER ARG ARG TYR ARG GLY ASP PHE ALA GLU MET SER SER GLN LEU HIS LEU THR PRO  
 TTC TOC CCG CCG TAC CCG GCG GAC TTC GCG GAG ATG TOC ACG CAG CTG CAC CTG ACG CCG

400

PHE THR ALA ARG GLY ARG PHE ALA THR VAL VAL GLU GLU LEU PHE ARG ASP GLY VAL ASN  
 TTC ACC GCG CCG GGA CCG TTT GCG ACG GTG GTG GAG GAG CTC TTC ACG GAC GCG GTG AAC

TRP GLY ARG ILE VAL ALA PHE PHE GLU PHE GLY GLY VAL MET CYS VAL GLU SER VAL ASN  
 TGG GCG AGG ATT GTG GCG TTC TTT GAG TTC GGT GCG GTC ATG TGT GTG GAG ACG GTC AAC

500

ARG GLU MET SER PRO LEU VAL ASP ASN ILE ALA LEU TRP MET THR GLU TYR LEU ASN ARG  
 CCG GAG ATG TOG CCG CTG GTG GAC AAC ATC GCG CTG TGG ATG ACT GAG TAC CTG AAC CCG

600

HIS LEU HIS THR TRP ILE GLN ASP ASN GLY GLY TRP VAL GLY ALA SER GLY ASP VAL SER  
 CAC CTG CAC ACC TGC ATC CAG GAT AAC GGA GCG TGG GTA GGT GCA TCT GGT GAT GTG AGT

▲

LEU GLY (000) GGCCACAGGTCCGAGATGGGGGTTGAGTGGGGTGGGCTOCTGGCCAATGGAGGCTGTGGAGCC  
 CTG GCG TGA

FIG. 3